









## Recommendations for achieving interoperable and shareable medical data in the USA

Ana Szarfman <sup>1✉</sup>, Jonathan G. Levine <sup>2</sup>, Joseph M. Topping <sup>3</sup>,  
Frank Weichold <sup>1</sup>, John C. Bloom <sup>4</sup>, Janice M. Soreth<sup>5</sup>,  
Mark Geanacopoulos<sup>1</sup>, Lawrence Callahan <sup>1</sup>, Matthew Spotnitz <sup>6</sup>, Qin Ryan<sup>1</sup>,  
Meg Pease-Fye <sup>1</sup>, John S. Brownstein <sup>7</sup>, W. Ed Hammond <sup>8</sup>,  
Christian Reich <sup>9</sup> & Russ B. Altman <sup>10</sup>

### Abstract

Easy access to large quantities of accurate health data is required to understand medical and scientific information in real-time; evaluate public health measures before, during, and after times of crisis; and prevent medical errors. Introducing a system in the USA that allows for efficient access to such health data and ensures auditability of data facts, while avoiding data silos, will require fundamental changes in current practices. Here, we recommend the implementation of standardized data collection and transmission systems, universal identifiers for individual patients and end users, a reference standard infrastructure to support calibration and integration of laboratory results from equivalent tests, and modernized working practices. Requiring comprehensive and binding standards, rather than incentivizing voluntary and often piecemeal efforts for data exchange, will allow us to achieve the analytical information environment that patients need.

Reported world-wide mortality from COVID-19 has surpassed 6 million with over 16% of deaths in the USA alone<sup>1</sup>. Despite our vaccination efforts against COVID-19, the analytical deficiencies of USA health information systems (HIS) uncovered by the pandemic remain largely unresolved<sup>2</sup>. We still cannot answer basic questions that should be answerable by a simple query of the data, such as, what is the mortality rate according to patient variables? Also, public health systems and practitioners are still forced to rely on outmoded forms of communication (e.g., paper and fax) which do not provide rapid access to needed information.

Although recognized as a leader in advancing cutting edge biomedical research and medical technology, the USA continues to rely on multiple, independent healthcare systems and versions that cannot seamlessly communicate with each other. This lack of interoperability within and across hospital systems, laboratories, public health programs, physicians' offices, and regulatory and research data resources hinders rapid improvements in medical treatment, public health, decision-making, and research. The main reason for the failure to achieve interoperability and for the information loss, inefficient operations, and huge (and frequently hidden) costs that result, is the lack of comprehensive, centrally coordinated, fully validated, traceable, and enforceable medical data collection and transmission standards<sup>3</sup>. Bi- and

<sup>1</sup>U.S. Food and Drug Administration, Silver Spring, MD 20993, USA. <sup>2</sup>Independent Researcher, Rockville, MD 20855, USA. <sup>3</sup>Your Health Concierge, Inc., Bethesda, MD 20815, USA. <sup>4</sup>College of Veterinary Medicine, Purdue University, West Lafayette, IN 47907, USA. <sup>5</sup>Independent Medical Consultant, Chevy Chase 20815 MD, USA. <sup>6</sup>Columbia University Irving Medical Center, New York, NY 10032, USA. <sup>7</sup>Harvard Medical School, Boston Children's Hospital, Boston, MA 02115, USA. <sup>8</sup>Duke Center for Health Informatics, Duke Clinical & Translational Science Institute, Durham, NC 27705, USA. <sup>9</sup>Columbia University Department of Biomedical Informatics, New York, NY 10032, USA. <sup>10</sup>Departments of Bioengineering, Genetics & Medicine, Stanford University, Stanford University School of Medicine, Stanford, CA 94305-4145, USA. ✉email: [ana.szarfman@fda.hhs.gov](mailto:ana.szarfman@fda.hhs.gov)

multi-directional feedback loops that are needed for prompt access to ancillary data, for clarifications, and for quickly reporting and addressing system and data errors are also lacking. Without easy access to this additional information, electronic health records (EHRs) cannot be made portable<sup>4</sup>, and the full potential of these records to support research and innovation cannot be realized.

Data that can be easily exchanged is a goal that many parties have long been advocating for. The COVID-19 pandemic has made this issue more urgent than before as we have “*to move faster than the virus*” [Personal communication from Dr. Mirta Roses]. Unfortunately, to date the COVID-19 pandemic has only underscored the consequences of having information systems that rely on non-binding standards for data management and exchange, standards that are themselves based on multiple, unreconciled data models. In principle, a data model should provide universal definitions of data elements (i.e., units of data having a precise meaning and interpretation) for users of heterogeneous data sites that want to share or aggregate data, to allow them to speak a common language<sup>5</sup>.

Emerging technologies that promise to revolutionize healthcare add additional urgency to efforts to achieve interoperability in our HIS. In a decade, there will be more sources of data<sup>6</sup>, for example from wearable devices such as the Apple Watch and Fitbit, that patients will use to record information. By combining these data with artificial intelligence and machine learning, in which machines are able to automatically process the data, diagnosis and prediction of patient outcomes could be improved. However, the successful use of these computational tools strongly depends on accurate collection and exchange of massive amounts of complex data derived from next-generation sequencing, imaging devices, laboratory assays, and many other sources. Unfortunately, the data being collected remain predominantly in local silos and frequently the data are neither standardized nor of the quality required by these advanced automated approaches<sup>7,8</sup>. The promise of all of these technological and scientific advances will be unrealized without interoperable standards that are fully representative of real-world clinical data (not just based on theoretical examples), and are fit-for purpose for data collection, exchange, integration, and analysis, and traceable to the original information.

Perhaps the most challenging roadblock for implementing interoperability for data collection is the tolerance for highly customized, proprietary HIS and their unique versions. The inconsistencies created by unnecessary customization creates a state of confusion that makes it impossible to reliably identify in a timely fashion critical data facts and inconsistencies that must be communicated to those making critical decisions about how these systems should be designed and implemented. These decision makers include those in government organizations, the vendors of HIS, software developers, and other stakeholders, including patients and patient advocates.

By presenting the following description of deficiencies in the USA health information system and recommendations for addressing them at their root causes, we hope to stimulate constructive dialog among multiple stakeholders and inform policy changes in the USA and other countries where such measures are needed. Recognizing our ethical responsibility to rapidly provide the best information to help patients<sup>9</sup>, we propose building an alternative, more transparent system based on interoperability that starts at the data collection stage. This alternative system would be one in which the benefits of new computational technologies can be realized, where patients are able to take control of their data, and where accurate and timely data can be rapidly shared to advance medical research and improve public health.

### **The lack of universal and harmonized data collection and transmission standards**

To date, policies to increase interoperability in our HIS have been based on downstream transactions, for example they seek to improve e-prescribing, billing, health information exchange, certification of EHRs<sup>10,11</sup>, and regulatory submissions. These policies do not enforce a universal standard for the collection and transmission of defined variables and values for each data element, even for straightforward information such as demographic data<sup>5</sup>. Lacking universal standards, most health data exchange is therefore subject to the custom constraints of a multitude of unique, proprietary HIS, and the non-interoperable, disparate versions of the data elements in these systems. For example, proprietary HIS systems modify most lab data received in their databases by mapping them to built-in terms. This results in a multitude of data conversion cycles that are difficult to document and untangle, because they are not traceable to the original data elements. Mapping and remapping from the irregular internal codes of each HIS version to the standardized versions needed for exchanging data is an error-prone, inefficient, and costly process that is repeated in reverse at the receiving end(s) when integrating the exchanged data back into the internal codes of the HIS version in which they were received.

There are ongoing efforts by the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services in the U.S. to support data exchange via secure FHIR HL7 application programming interfaces<sup>12</sup> (i.e., software exchange engines created by The Health Level Seven International healthcare standards organization). However, without common, enforceable, and well-documented data structure and coding across pertinent HIS and application programming interfaces, data exchange may still require manual, and frequently blinded mapping, which makes rapid transfer of information unfeasible.

To achieve a health system that enables continuous improvements, we need systems that collect the data that are most important for patient care, for accomplishing critical analyses, for enhancing the level of evidence, and for addressing public health challenges<sup>13–16</sup>. Therefore, we must focus on developing universal standards for the collection and validation of the most clinically important data as they are created (e.g., results from centrally calibrated laboratory tests during the entire course of clinical care). Only when such standards are in place can we ensure that valid information is being correctly captured and delivered. We must also ensure that the diverse software, transfer engines, and information technology systems can correctly interpret these standards, and process standard nomenclatures and notations without corruption. Redundant backup systems, feedback loops for prompt and early identification and communication of problems, and automated data verification processes will be needed to ensure data integrity and identify and correct the sources of transmission errors. Options should be provided for the public to monitor the accuracy of their medical data throughout all encounters (e.g., prescriptions, diagnosis, procedures), in the same way they can monitor their interactions with the Social Security System or banking institutions.

A recent collaborative effort between HL7 International, which provides common standards for exchange of data in healthcare, and the Observational Health Data Sciences and Informatics (OHDSI) collaborative, which defines and maintains the common data model known as OMOP for international observational research studies, seeks to implement a unique data model for assembling and sharing information gathered in clinical care. This undertaking should enable us to integrate clinical data within huge repositories for advanced analytics, without the information loss caused by sequential mapping and remapping

from and to a multitude of untraceable data models<sup>17</sup>. However, HL7 and OHDSI are not providing interoperable standards for the collection of factual data into EHRs. Without strong legislative support, funding, and enforcement, an interoperable model for data collection at the source that can fully address our most critical health information needs will not become a reality.

Although recommendations for addressing data interoperability in our HIS are described in the policy documents of organizations involved in oversight<sup>11</sup> and in the scientific literature<sup>18</sup>, much of the medical and scientific community remains insufficiently aware of the limitations of these systems, and tackling our widespread usability problems has not become a universally shared priority. We suggest that recent failures in disease prediction models<sup>19–21</sup> can be attributed in part to irregularities in how data are captured, exchanged, and maintained, and our inability to systematically access and compare these data across multiple EHR systems and versions over time.

To build quality data systems, we must have reliable enforcement mechanisms in place to monitor the implementation of and adherence to interoperable data standards. To monitor such process, we need to conduct Good Clinical Practice inspections and adopt reliable monitoring tools and enforcement mechanisms (analogous to those used by the Treasury Department to assure honesty of monetary transactions). These inspections will require highly trained professionals capable of detecting inaccurate data, improper coding, and failures of prediction models that clinicians rely on.

There are limits to the time healthcare professionals can (and should) spend entering data. A core principle of informatics is that data should only be entered once, and whenever possible by the device collecting the data. In the scenarios where automated entry is not possible, an interoperable system should facilitate data entry and coding by providing automated, interactive graphic representations of the data already in the system and smarter options with standard terminology for outcomes for given symptoms, diseases, medications, and patient profiles. Establishing high quality hardware and software systems for collecting and delivering interoperable and fully traceable healthcare data to users would also create a dynamic in which it would be easier to assess the value and cost of data, and what additional data should be captured. Furthermore, the creation and support with reimbursable billing codes of large numbers of positions for scientifically trained clinical information professionals to manage medical information systems will increase the value of these systems for caregivers, researchers, and patients.

### Issues requiring prompt attention

**Lack of ascertainment of unique patients.** Although HIPAA initially required the creation of a health identifier in 1996<sup>22</sup>, federal funds for unique universal patient identifiers have been banned since Congress prohibited their use due to privacy concerns<sup>23</sup>. Our failure to implement national, unique identifiers in the USA linking a patient's data to their healthcare professionals and HIS systems leads to unlinked, incomplete, and often duplicated records, and is another significant source of data quality problems that have been avoided in countries that have implemented unique identifiers<sup>23</sup>. In addition, it is still nearly impossible for a person to access their own vaccination records if they are in databases separate from their EHR records or were submitted by paper or fax. It is also difficult or impossible to carry out the early cancer prevention studies<sup>24</sup> that require that complete clinical information be linked to the correct patients even when they change health providers.

Although the prospect of unique patient identifiers raises valid privacy concerns, it can be argued that it would be easier to

monitor and protect privacy with a single, properly encoded universal identifier than with a multitude of poorly documented ones. The absence of a unique identifier is actually one of the greatest causes of invasion of privacy, because typically over half of the EHRs in an institution will mistakenly include someone else's data [Personal communication by Dr. W. Ed Hammond] that may be identifiable.

The current reliance on data aggregation techniques to protect patient privacy significantly delays our access to the information and impedes our understanding of the trajectory of diseases in individual patients, with potentially adverse consequences for their medical care and for identifying critical patient-level variables for subsequent research studies. We must therefore invest in better and updated privacy protection systems and law enforcement solutions. As data scientists, we are concerned about the limitations of HIPAA for privacy protection, due to the ease that such data can be re-identified. Our laws and regulations need to balance individual privacy protection, with making data available for improving health outcomes. At a minimum, the approach to governance we adopt must ensure the following: the system is able to identify and control who can have the authorized level of access to the medical records; every user has a unique ID and a secure password; audit trails are used to track every user activity, and to provide accountability; only authorized personnel can access audit trails, and assess who has accessed or modified a record; and the data storage provider is not able to access personal identifiable information.

A single patient identifier also has health equity ramifications in its favor. Patients who are poorer typically have less insurance coverage or none at all and often switch healthcare systems. They are underrepresented in HIS and research studies, and less likely to have their specific needs understood. A unique identifier should improve the representation of these patients in our HIS and thus our ability to address health inequities.

**Lack of information about patient mortality.** The inadequacy of our current system for data collection is well illustrated by our failure to collect data as fundamental as mortality in a standardized fashion. Fatal outcomes are not incorporated into the medical record unless death occurs during hospitalization. When needed for public health measures, epidemiological studies, and other research, data on death may be obtained from private services that collect information from funeral homes and obituaries, disease registries unconnected to EHRs, or from the National Death Index website. This website is typically late in gathering mortality information as it is collected by a multitude of disparate local and state systems before being reported to the National Center for Health Statistics. Comprehensive data on mortality and cause of death should be methodically linked to clinical data for the over 330 million individuals in the USA (as we have begun to do for COVID-19 cases). This information will allow for the creation of focused decision support systems for clinical data that are better designed to prevent serious and fatal medical errors, one of the top causes of death in hospitals in the USA<sup>25</sup>.

**Poorly codified and calibrated clinical laboratory data.** Clinical laboratories began collecting digitized data in the 1960s. Although these data support 60 to 70 percent of decisions related to diagnosis, treatment, hospital admission, and discharge, they remain poorly codified, complicated to process, and are underused for medical decision-making and research.

USA programs that defined the minimum government standards for EHRs have offered laboratories incentives to adopt proposed standards for messaging and encoding laboratory data.

Unfortunately, serious functional problems still exist with the coding of laboratory test identifiers. There are multiple ways for the same analytes to be represented by different labs and instruments and this results in improper assessments of coded terms and incorrect code selection and categorization. Moreover, coding systems often do not allow for transparent incorporation and transmission of the limits of detection of a test, the presence of interfering substances, and how a particular analyte is measured. Also, failure to enforce the use of consistent quantitative units of measure is a frequent source of data errors.

There is a pressing need for an expanded infrastructure to support the collection and distribution of the stable reference standards needed to support the accurate calibration and safe integration of the results from equivalent tests measuring the same analyte, performed by different instrument platforms or laboratories<sup>26,27</sup>. The Office of the National Coordinator for Health Information Technology recognizes this problem when it states, “Harmonization status indicates calibration equivalencies of tests and is required to verify clinical interoperability of results. Tests that are harmonized may be interpreted and trended together, and may use the same calculations, decision support rules, and machine learning models. Tests that are not harmonized should be interpreted and processed individually, not in aggregate with other tests.”<sup>3</sup>

This infrastructure will simplify the identification of a natural functional interoperability pathway that can be used as a backbone for integrating the currently unwieldy, inconsistent, and incomplete data coding standards for laboratory data. An illustration of the consequences of the failure to fully standardize laboratory data collection and calibration of the results is the limited understanding of the evolving prevalence of COVID-19, due to the inability to account for the performance differences of the over 1,000 SARS-CoV-2 diagnostics that are listed worldwide<sup>28</sup>. We also need to understand their performance characteristics according to the particular purpose for which a test is being performed (e.g., permission to travel, to access specific facilities, etc.)<sup>29</sup>.

**Business practices that hinder modernization.** The world-wide-web and online business transaction systems such as Amazon’s e-commerce system were built with a clear understanding of the value of interoperability. These systems ensure that the correct data are collected and stored in an organized, automatically aligned format that is optimized to address new communication requirements and analytical functions. Realizing this scenario for health data will require changes in current practices. Since individual enterprises have built one-of-a-kind systems, there are often strong financial reasons not to share proprietary information. Current laws prohibiting information blocking have not accomplished their purpose, because it is impossible to effectively oversee the thousands of unique versions of HIS.

Given this scenario, it would be useful, once the needed information and data routes are identified and categorized, to develop prototype systems to demonstrate the benefits of profound change in how we manage health information. The development, testing, and validation of these prototypes for addressing the various requirements of patient care and research and development should be based on the integrity, completeness, traceability, and usability of the data; on the avoidance of preventable medical errors; and on measurable improvements in health outcomes.

**Improving the processing of laboratory data linked to the regulatory activities of the FDA.** In contrast to other data transactions for which federal regulations are seeking to increase

interoperability (e.g., using ICD-10 coding for billing), in the USA there is no clear business model that incentivizes standardization of laboratory data coding and its integration across medical encounters. Nor is there a single coordinated authority in the USA to monitor and enforce the adoption of, and adherence to, such standards or the transmission of intact laboratory data to end users. Interoperable standards for laboratory data are still very immature (paper and fax lab submissions are still commonplace), and still rely on billing codes for managing and understanding this information, despite their limited scope. For example, there are only 12 Current Procedural Terminology codes used for billing reimbursement that identify the COVID-19 or SARS-COV-2 infectious agent or their antibody response<sup>30</sup>, while the FDA lists 357 identifiers for COVID-19 testing<sup>31</sup>.

Therefore, we suggest that one area that we should use as a model for how to achieve interoperability of patient data, and where favorable incentives for reform may already exist, is in the processing of clinical laboratory data in drug marketing applications submitted to the FDA. Currently, such data undergo multiple transformation steps before regulatory submission, and although results in a given new drug application may be calibrated, the results for many equivalent analytes coming from different sponsors, laboratories, and instruments are not necessarily calibrated the same way<sup>3,26,27</sup>.

We propose to begin the process of prototype development by creating a centralized calibration process for routine and critical analytes so that results collected during clinical trials will be equivalent regardless of the instrument or the laboratory. The aim is to eliminate the severe problems that result from customized data systems and demonstrate that time-consuming mapping and translation errors, and the associated loss of information, can be avoided while adding traceability and clarity to the clinical laboratory data in marketing applications. The recent phenomenon of increased mergers between central labs supporting pharmaceutical company sponsors and labs that support hospital networks will enable the systematic identification and removal of many deficiencies that derived from multiple sources of lab data, and help implementation of robust and universal data standards. We expect that the time and cost savings and the gains in accuracy demonstrated by a prototype system for clinical laboratory data will be welcomed by the pharmaceutical and device industries, the research and public health communities, and patients. In its processing of lab data, this initiative will include all the standardized data elements needed for analysis of regulatory data submissions, including those related to demographics, diagnosis, medical history, laboratory tests, death, and cause of death. Such standards will greatly enhance regulatory review of marketing applications across multiple sponsors and facilitate comparison of clinical trial lab results across applications, providing valuable feedback to the pharma sponsors.

When it reaches a level of maturity, the prototype for handling laboratory and other clinical data in regulatory submissions could be expanded to non-regulatory contexts, including routine patient care. The lessons learned could eventually be applied to the evaluation and certification of EHRs and decision support systems. The knowledge gained in how to create a truly interoperable system could also be used to address the analytical needs of other data resources including registries, repositories of real-world data, and regional data exchanges.

#### Returns on investment

Adoption of our recommendations will simplify the continual enhancement, maintenance, oversight and the analytical

**Box 1: | Recommended steps for legislative action**

- Empower an oversight and enforcement agency with a qualified advisory board representing all stakeholders to identify and address critical usability requirements for building an interoperable & interconnected Health Information System in the USA
- Enforce the creation and maintenance of a thorough common data model for clinical data
- Prohibit unwieldy data customization by enforcing interoperable and interconnected standards for medical data collection for every organization that collects or processes medical data (e.g., hospitals, laboratory information systems)
- Establish automated data verification processes to confirm that the data collected are transmitted without distortion to correct patient records and end users; identify problems through feedback loops, and correct the sources of any data errors
- Enforce a standard, certifiable calibration process that ensures that different tests for a given analyte give equivalent results regardless of the instrument used or the laboratory performing the test
- Implement a universal, unique patient identifier secured by the strongest privacy-enhancing technology and supported by a security infrastructure
- Authorize a central body to collect death and cause of death information for all individuals in the USA, with the federal government defining the requirements and precautions needed to avoid fraud
- Require the Centers for Medicare & Medicaid Services to create reimbursable billing codes for clinical informatics professionals who can make informed decisions about HIS selection, optimization of analytical and decision support functions, and maintenance
- Establish Good Clinical Practices with adequate inspections of the analytical clinical data processes and facilities
- Create incentives aligned with patients' and public health needs in which healthcare vendors are rewarded for documenting and avoiding medical errors and unnecessary processing costs
- Identify and correct gaps and inconsistencies in current regulatory requirements

**Box 2: | Recommended steps for public-private partnership action**

- Catalog the information required for building the decision support systems needed to improve the quality of patient care, and maintain updated information
- Authorize an advisory board representing all stakeholders, subject matter experts, and patient advocates to identify and address current roadblocks for collecting and transmitting interoperable clinical data in a fully traceable manner
- Work with Standards Development Organizations to determine how best to codify clinical data at the point of data collection, and maintain continuous quality improvements in coding
- Identify the shortest pathways for information to reach the correct patient record and authorized end users
- Adopt data collection systems that are fully traceable to the original data facts and thus make it possible to locate the sources of medical and system errors to avoid their recurrence
- Identify and link important health information data that are currently not connected to hospitals, including death, cause of death, data collected by registries, and vaccinations
- Implement a pilot system for collection of standardized, calibrated clinical laboratory data, and ancillary information, starting with the clinical data submitted to the FDA by drug companies
- Assess progress based on measurable improvements in data integrity and completeness, analytics, healthcare delivery, and auditability, as well as reduced operating costs
- Make the data elements and documentation of the mature prototype(s) available in a public repository for testing and feedback from users

functions of a fully interoperable, public health and medical data system. The savings achieved through interoperability across the Research and Development value chain would expedite the discovery and development of safe and effective vaccines, treatments, and the identification of marketed drugs that can be repurposed to treat patients. Analysts will be able to discover consistent and reproducible efficacy and safety signals within and across multiple data resources and perform meta-analyses of all selected data rather than focusing on limited, static summary reports<sup>32,33</sup>. Automated analytical tools will remain securely linked to the original data, making it possible to quickly complete additional evaluations of emerging issues. In support of these predictions, countries that have more interconnected HIS have been able to analyze their medical data more efficiently, and have provided important findings. For example, the rapid completion of the dexamethasone study in the United Kingdom in patients with COVID-19<sup>34,35</sup> would have been very difficult to achieve in the USA with our highly customized and uncoordinated systems for capturing patient-level data. Also, Israel with its standardized, highly interoperable medical information system and a universal patient identifier has provided critical information about

breakthrough infections in patients who were considered to be fully vaccinated<sup>36,37</sup>.

In a fully interoperable health information system, patients will receive improved medical care based on the ability of clinicians to detect, troubleshoot, and prevent critical and costly medical and system errors and benefit from public health measures that are based on information that is reliable, complete, and up-to-date. Although the deconstruction and rebuilding that we are proposing may be costly, it will be even more expensive to continue to undertake never-ending customization and processing of data to fit the unpredictable, continuously changing constraints of multiple, incompatible silos. To generate popular support for transforming our HIS, we must inform the public of the risks of medical errors associated with a failure to accurately transmit critical information to the correct patient record and the risks for data quality posed by the current methods of maintaining data confidentially and privacy. Once we achieve true interoperability, it will become obvious that all health data are important, and we are ethically bound to adopt data retention policies that will preserve this information for our needs and for future generations.

## Conclusions

In conclusion, the COVID-19 crisis is another wake-up call that reminds us that we cannot continue to use outdated data solutions that jeopardize our ability to advance research capabilities<sup>4</sup>, and can lead to medical errors and loss of life. Boxes 1 and 2 offer a set of summary recommendations that, if adopted, would help achieve needed solutions to the problems being described in this paper.

If Amazon can track packages, international banks can track money, and weather maps can track complex weather patterns, we can also learn how to track and analyze complex health data. Our recommendations are intended to create the conditions in which we can address an entrenched and highly complex problem that will only become worse if unaddressed. This problem will not be cheap to fix, but it will be much costlier to ignore.

Received: 12 October 2021; Accepted: 20 June 2022;

Published online: 18 July 2022

## References

- WHO. *Worldometer Coronavirus*. <https://www.worldometers.info/coronavirus/#countries> (2022).
- NCB News. *Government Watchdog Says Key Federal Health Agency is Failing on Crises*. <https://www.nbcnews.com/politics/politics-news/government-watchdog-says-key-federal-health-agency-failing-crises-n1288154> (2022).
- ISA. *The Office of the National Coordinator for Health Information Technology. 2022 Interoperability Standards Advisory*. <https://www.healthit.gov/isa/sites/isa/files/inline-files/2022-ISA-Reference-Edition.pdf> (2022).
- Denny, J. C. & Collins, F. S. Precision medicine in 2030—seven ways to transform healthcare. *Cell* **184**, 1415–1419 (2021).
- U.S. FOOD & DRUG. *COVID-19 Real World Data (RWD) Data Elements Harmonization Project*. <https://www.fda.gov/drugs/coronavirus-covid-19-drugs/covid-19-real-world-data-rwd-data-elements-harmonization-project/> (2022).
- AWS Data Lake Team. *A Public Data Lake for Analysis of COVID-19 Data by AWS Data Lake Team 08 APR 2020*. <https://aws.amazon.com/blogs/big-data/a-public-data-lake-for-analysis-of-covid-19-data/> (2022).
- Comstock, J. *ONC, CDC Want to Fix the Fragmented Public Health System COVID-19 Exposed*. <https://www.healthcareitnews.com/news/onc-cdc-want-fix-fragmented-public-health-system-covid-19-exposed> (2021).
- Achenbach, J. & Abutaleb, Y. *Messy, Incomplete U.S. Data Hobbles Pandemic Response*. September 30, 2021 at 9:30a.m. EDT. <https://www.washingtonpost.com/health/2021/09/30/inadequate-us-data-pandemic-response/> (2021).
- Montague, E. et al. The case for information fiduciaries: the implementation of a data ethics checklist at Seattle children's hospital. *J. Am. Med. Inform. Assoc.* **28**, 650–652 (2021).
- CMS. *Interoperability and Patient Access Fact Sheet*. Mar 09, 2020. <https://www.cms.gov/newsroom/fact-sheets/interoperability-and-patient-access-fact-sheet/> (2020).
- CMS. *Medicare and Medicaid Promoting Interoperability Program Basics*. <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Basics> (2022).
- Frieden J. *MedPage Today*. May 6, 2021. *Q&A: Talking Health IT With Micky Tripathi*. <https://www.medpagetoday.com/practicemanagement/informationtechnology/92459> (2021).
- Institute of Medicine (USA). *The Learning Healthcare System: Workshop Summary* (National Academies Press, 2007).
- Hartley, D. M. & Seid, M. Collaborative learning health systems: Science and practice. *Learn. Health Syst.* **5**, e10286–e10286 (2021).
- Parsons, A. et al. Seven practices for pursuing equity through learning health systems: notes from the field. *Learn. Health Syst.* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8278437/> (2021).
- Ros, F. et al. Addressing the Covid-19 pandemic and future public health challenges through global collaboration and a data-driven systems approach. *Learn. Health Syst.* <https://doi.org/10.1002/lrh2.10253> (2021).
- OHDSI. *HL7 International and OHDSI Announce Collaboration to Provide Single Common Data Model for Sharing Information in Clinical Care and Observational Research Leading Organizations Will Integrate Products to Create a Single Source for the Sharing and Tracking of Data*. [http://www.hl7.org/documentcenter/public/pressreleases/HL7\\_PRESS\\_20210301.pdf](http://www.hl7.org/documentcenter/public/pressreleases/HL7_PRESS_20210301.pdf) (2021).
- Conway, J. R., Warner, J. L., Rubinstein, W. S. & Miller, R. S. Next-generation sequencing and the clinical oncology workflow: data challenges, proposed solutions, and a call to action. *JCO Precis.Oncol.* <https://doi.org/10.1200/po.19.00232> (2019).
- Wong, A. et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern. Med.* **181**, 1065–1070 (2021).
- Habib, A. R., Lin, A. L. & Grant, R. W. The epic sepsis model falls short—the importance of external validation. *JAMA Intern. Med.* **181**, 1040–1041 (2021).
- Ross C. *Epic's AI Algorithms, Shielded From Scrutiny by a Corporate Firewall, are Delivering Inaccurate Information on Seriously Ill Patients*. <https://www.statnews.com/2021/07/26/epic-hospital-algorithms-sepsis-investigation/> (2021).
- Public Health Law. *Health Insurance Portability and Accountability Act of 1996*. <https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf> (1996).
- Sood, H. S., Bates, D. W., Halamka, J. D. & Sheikh, A. Has the time come for a unique patient identifier for the US? *NEJM Catalyst* <https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0252> (2018).
- Friends of Cancer Research. *Friends of Cancer Research Virtual Meeting—A Path for Early Detection*. Streamed live on Mar 29, 2022 [https://www.youtube.com/watch?v=J\\_G0C7vN724](https://www.youtube.com/watch?v=J_G0C7vN724) (2022).
- Institute of Medicine (US) Committee on Quality of Health Care. *To Err is Human: Building a Safer Health System* (National Academies Press, USA, 2000).
- Paxton, A. *New hope for Lab Data Interoperability*. *CAP Today* 35.11. <https://www.captodayonline.com/new-hope-for-lab-data-interoperability/> (2021).
- Levenson, D. *Untangling Laboratory Data's Twisted Journey*. <https://www.aacc.org/cln/articles/2021/december/untangling-laboratory-datas-twisted-journey>. (2021).
- SARS-CoV-2 Diagnostic Pipeline. <https://www.finddx.org/covid-19/pipeline/> (2022).
- World Health Organization. *Statement on the Tenth Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Coronavirus Disease (COVID-19) Pandemic*. [https://www.who.int/news/item/19-01-2022-statement-on-the-tenth-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-coronavirus-disease-\(covid-19\)-pandemic](https://www.who.int/news/item/19-01-2022-statement-on-the-tenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic) (2022).
- AMA. *COVID-19 CPT Coding and Guidance*. <https://www.ama-assn.org/search?search=cpt-emergency-release-covid-related-code-file.xlsx> (2022).
- U. S. FOOD & DRUG. *COVID-19 Tests and Collection Kits Authorized by the FDA: Infographic*. <https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/covid-19-tests-and-collection-kits-authorized-fda-infographic> (2021).
- Janet Woodcock, M.D., Amy Abernethy, M.D. *FDA's Data Modernization Action Plan: Putting Data to Work for Public Health*. <https://www.fda.gov/news-events/fda-voices/fdas-data-modernization-action-plan-putting-data-work-public-health/> (2021).
- Announcement: Towards greater reproducibility for life-sciences research in nature. *Nature*. *Nature* **546**, 8–8 (2017).
- The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. *N. Engl. J. Med.* **384**, 693–704 (2021).
- NHS digital. *Spine* <https://digital.nhs.uk/services/spine> (2021).
- Bar-On, Y. M. et al. Protection of BNT162b2 vaccine booster against Covid-19 in Israel. *N. Engl. J. Med.* **385**, 1393–1400 (2021).
- U. S. FOOD & DRUG. *Vaccines and Related Biological Products Advisory Committee September 17, 2021 Meeting*. <https://www.fda.gov/advisory-committees/advisory-committee-calendar/vaccines-and-related-biological-products-advisory-committee-september-17-2021-meeting-announcement> (2021).

## Acknowledgements

The views expressed in this article are those of the authors and do not necessarily represent the views or policies of the Food and Drug Administration or of the other institutions. The authors wish to acknowledge the valuable insights and discussions with Norman Stockbridge, Robert Temple, Mitra Rocca, Helena Sviglin, Frank Cucino, and Gregory Pappas from the U.S. Food and Drug Administration; Ingeborg Holt, informatician; Andrea Pitkus, Laboratory Informaticist and Clinical Terminology expert; Riki Merrick, Association of Public Laboratories; Sharonna Hoffman, Case Western Reserve University School of Law; Nanguneri Nirmala, Director, Center for Clinical Evidence Synthesis, Tufts Medical Center; Mirta Roses Periago, World Health Organization Special Advisor on COVID-19 for Latin America and the Caribbean and Sir George Alleyne, both Directors Emeritus of the Pan American Health Organization; and Sean Khozin, Chief Executive Officer, CancerLinQ and former Associate Director of the FDA Oncology Center of Excellence.

### Author contributions

A.S., J.G.L., J.M.T., F.W., J.C.B., J.M.S., M.G., L.C., M.P. designed the study, developed critical concepts, and wrote the paper; C.R., W.E.H., J.C.B., R.B.A. added clarity to critical concepts; M.S., Q.R. contributed to the work methodology; all authors read, edited, and approved the final version of the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43856-022-00148-x>.

**Correspondence** and requests for materials should be addressed to Ana Szarfman.

**Peer review information** *Communications Medicine* thanks Joe Ledsam, David Bates, Francesca Cerreta, Rebecca Kush and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022